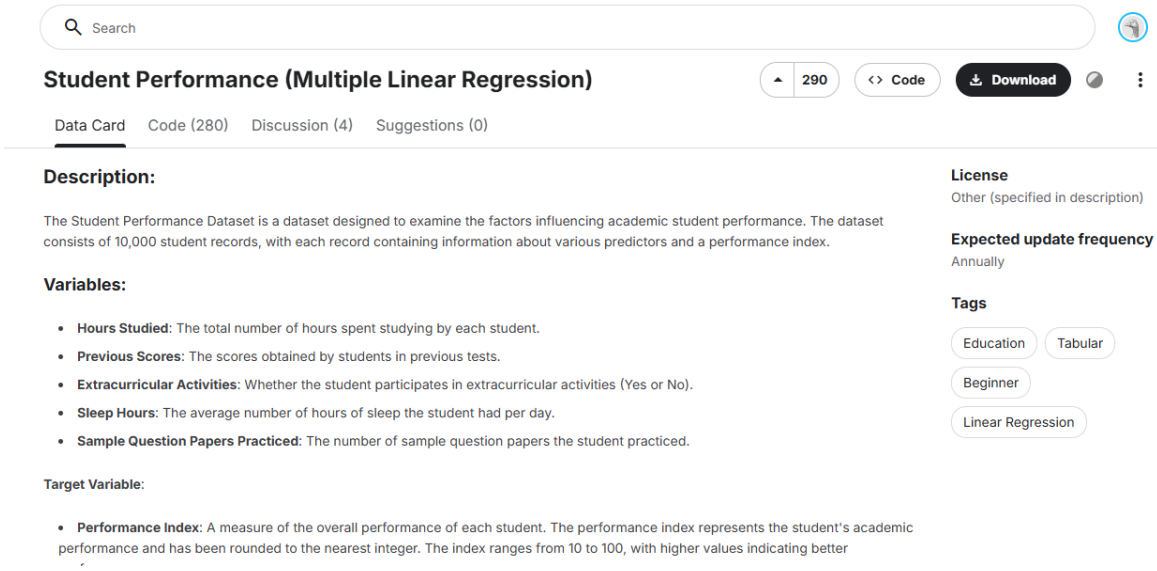
KELOMPOK:

Moh. Naufal Faqih 10222044

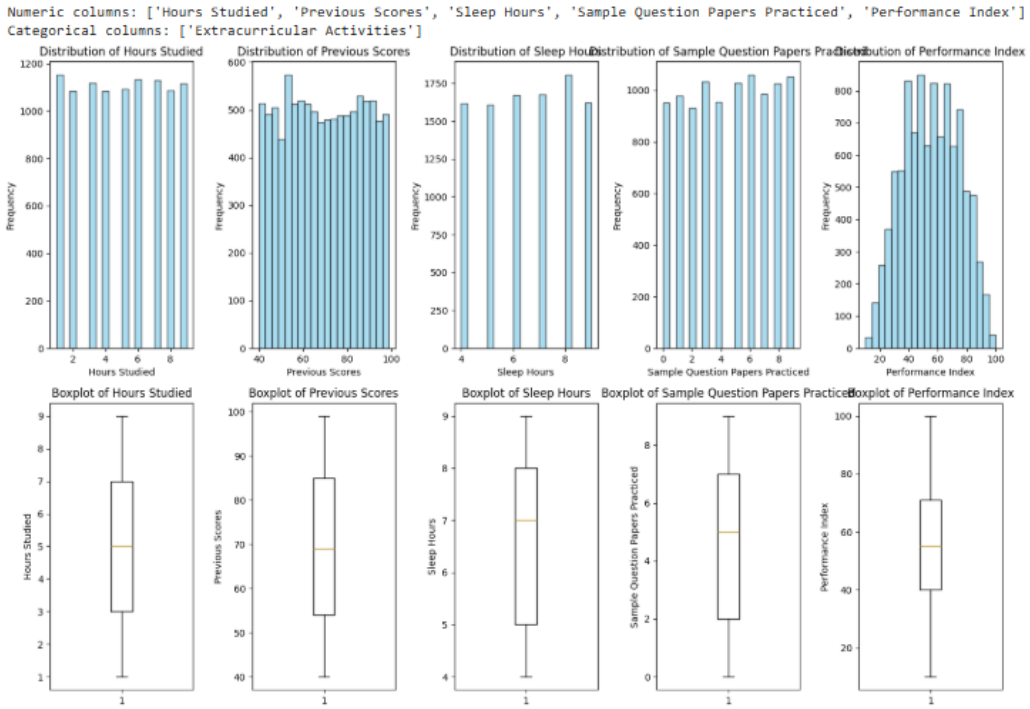
Firman Firdaus 10222033

EVALUASI MODEL REGRESI MULTIPLE LINEAR STUDI KASUS *STUDENT PERFORMANCE*

1. Dataset diambil dari: https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression



2. Mengidentifikasi Kolom Numerik dan Kategorikal



3. Melakukan Encoding Terhadap Nilai Kategorikal

```
Encoded Extracurricular Activities: {'No': np.int64(0), 'Yes': np.int64(1)}

Dataset shape after encoding: (10000, 6)

Encoded dataset info:
Hours Studied                        int64
Previous Scores                      int64
Extracurricular Activities           int64
Sleep Hours                          int64
Sample Question Papers Practiced     int64
Performance Index                  float64
dtype: object
```

4. Membagi Dataset Menjadi Data Training dan Testing

```
Features shape: (10000, 5)
Target shape: (10000,)

Feature columns: ['Hours Studied', 'Previous Scores', 'Extracurricular Activities', 'Sleep Hours', 'Sample Question Papers Pr
acticed']

Train set shape: (8000, 5)
Test set shape: (2000, 5)
```

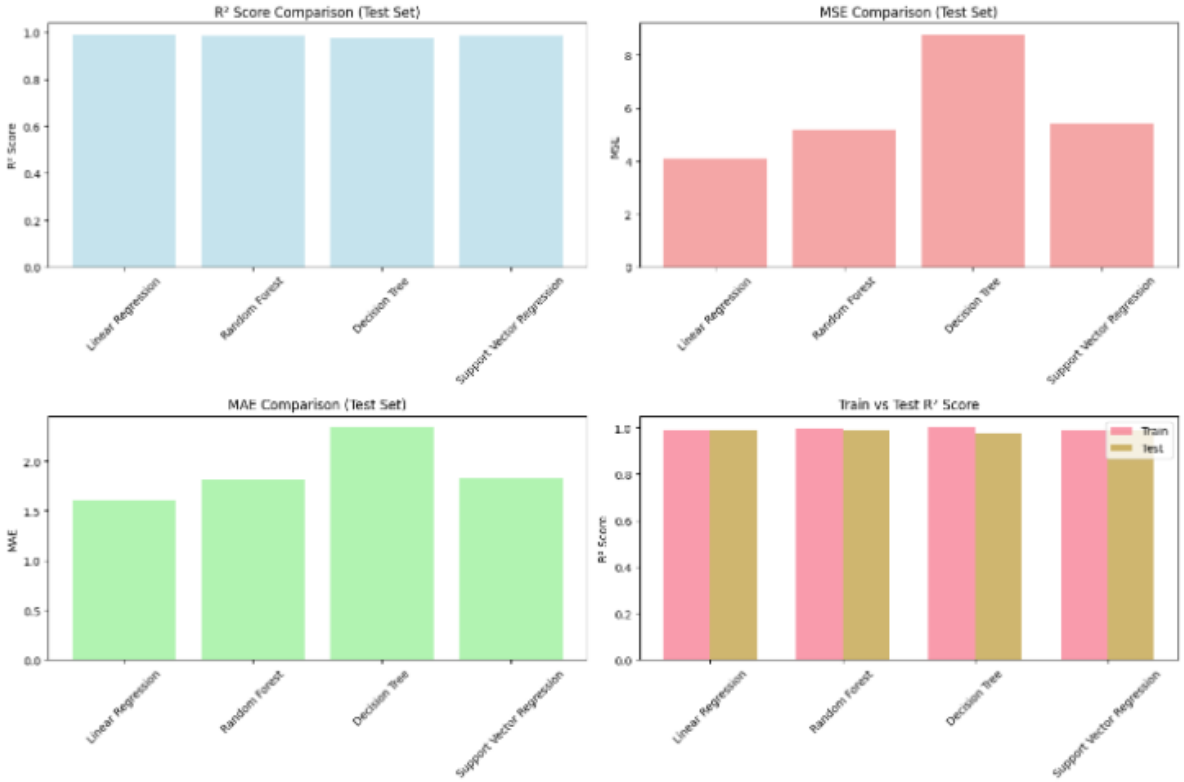
5. Menerapkan Feature Scalling

```
Feature scaling completed!
Feature means after scaling: [ 2.84217094e-17 -3.32178729e-16  5.15143483e-17  1.36779477e-16
 -8.17124146e-17]
Feature std after scaling: [1. 1. 1. 1. 1.]
```

6. Melatih dan Mengevaluasi Pada Beberapa Model

```
=== MODEL TRAINING & EVALUATION ===

Training Linear Regression...
Linear Regression Results:
  Train MSE: 4.1697
  Test MSE: 4.0826
  Train R²: 0.9887
  Test R²: 0.9890

Training Random Forest...
Random Forest Results:
  Train MSE: 0.9367
  Test MSE: 5.1719
  Train R²: 0.9975
  Test R²: 0.9860

Training Decision Tree...
Decision Tree Results:
  Train MSE: 0.2564
  Test MSE: 8.7701
  Train R²: 0.9993
  Test R²: 0.9763

Training Support Vector Regression...
Support Vector Regression Results:
  Train MSE: 5.2848
  Test MSE: 5.4206
  Train R²: 0.9857
  Test R²: 0.9854
```

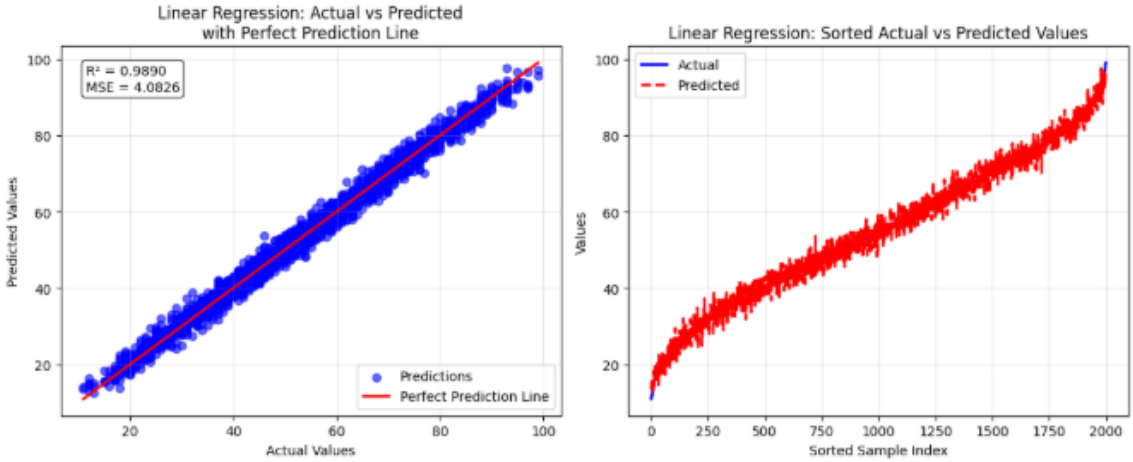
7. Perbandingan dari Berbagai Hasil Metrik Evaluasi Beberapa Model



8. Model dengan Hasil Evaluasi Terbaik

```
=== BEST MODEL ===
Best model: Linear Regression
Test R² Score: 0.9890
Test MSE: 4.0826
Test MAE: 1.6111
```

9. Visualisasi Data Hasil Evaluasi Model Terbaik



Kesimpulan

Model regresi linier berganda yang diterapkan pada dataset student performance menunjukkan hasil yang cukup baik dalam memprediksi nilai *Performance Index* berdasarkan beberapa fitur seperti jam belajar, nilai sebelumnya, aktivitas ekstrakurikuler, jam tidur, dan jumlah latihan soal. Evaluasi model dilakukan menggunakan metrik R^2 dan *Mean Squared Error* (MSE) pada data uji, di mana nilai R^2 menunjukkan bahwa model mampu menjelaskan proporsi variabilitas target dengan cukup baik. Visualisasi antara nilai aktual dan prediksi memperlihatkan bahwa hasil prediksi model cukup mendekati kenyataan. Selain itu, analisis koefisien menunjukkan pengaruh relatif masing-masing fitur terhadap performa akademik siswa, yang dapat dijadikan wawasan tambahan dalam pengambilan keputusan. Secara keseluruhan, model ini memberikan hasil yang memuaskan dan dapat dijadikan baseline yang baik sebelum mencoba pendekatan model prediktif yang lebih kompleks.